

Probability Models for DNA Sequence Evolution.

Rick DURRETT. New York: Springer-Verlag, 2002. ISBN 0-387-95435-X. viii + 240 pp. \$69.95 (H).

The DNA sequences of living organisms represent the outcome of a complicated stochastic process. Probabilistic models of this stochastic evolution, along with statistical methods for hypothesis testing and parameter estimation, are used to infer the nature of the history that gave rise to those DNA sequences.

DNA sequence evolution is generally a tree-like genealogical process, in which new variants arise, reproduce, and then either go extinct or reach *fixation*, the state in which individuals all descend from an initial mutant. This evolution operates on two main scales. Over shorter periods, such as the thousands of years of evolution of populations within a species, multiple variants coexist, and the salient features of the process are temporal changes in frequencies of these variants. Over longer periods, such as the millions of years that separate two species from their common ancestor, evolution can be viewed as a stochastic process of successive fixations of new mutations, or *substitutions*.

With numerous interspersed example datasets from the primary literature, this book offers a brief introduction to models of both long-term and short-term DNA sequence evolution, with emphasis on the latter. The fundamental approach in this field is as follows. To analyze DNA sequences from populations, a model of population demography—including migration patterns, time-varying sizes, and sex ratios—is selected. If the basic constant-sized model is used, then the genealogy of the sequences is treated as random using the *coalescent* model (Chap. 1); for complex demographies, an extension of the coalescent is used (Chap. 2). Mutations then follow a stochastic process, such as the *infinite-alleles* or *infinite-sites* model, conditional on the genealogy. In natural-selection models, unlike in selectively neutral models, the probability of survival of a lineage depends on its mutational type. Thus genealogy and mutation must be considered jointly, although approximations can help circumvent this difficulty (Chap. 3).

After exploring model properties, the book turns to hypothesis testing (Chap. 4). The tests of Tajima, Fu and Li, and Hudson, Kreitman, and Aguadé check whether genealogies within populations fit the predictions of the coalescent with infinite-sites mutation; a fourth test of McDonald and Kreitman considers whether sequence differences across species fit a neutral substitution model.

Diverse audiences will find much of value in this concise book. The author's rigor and ability to embed material in a broader mathematical context will appeal to quantitative readers with little background in biology. Given the brevity of the genetics introduction and the lack of biological motivation provided for most of the theory, however, such readers would do well to supplement the book with less theoretical treatments, such as those of Hartl and Clark (1998) or Li (1997). Geneticists will appreciate the assembly of details about familiar concepts into one place and the occasional corrections to original sources. However, the book is not mathematically self-contained; although much of the necessary background is in any introductory text on stochastic processes, some readers may be frustrated by an abundance of unreferenced mathematical facts.

The book's strength is its mathematical content, which revives almost-forgotten theorems, includes new proofs, and is thorough but not overwhelmingly detailed or overburdened with notation. The development often transits abruptly, however, so that the organization could be improved by conceptual overviews. The choice of material is somewhat quirky—for example, the treatment of the stepping-stone migration model is inexplicably lengthy and is only distantly connected to empirical data. Another unusual inclusion is the discussion of genome evolution in Chapter 5. Although this chapter makes better use of example datasets than do the earlier chapters, it is largely disjoint from the book's other topics.

Various omissions are worth noting. The book assumes that data come in the form of DNA sequences; although this type of data is in many ways ideal, models useful for other data types, such as DNA repeats, could be included. Recombination could be accorded space proportionate to the emphasis placed on it in current research. Although the tests in Chapter 4 are among those that are most popular, such issues as statistical power and the role of specific alternative hypotheses on the choice of test could be discussed. More important,

recent developments that move away from hypothesis tests toward model selection and maximum likelihood, Bayesian, and approximate Bayesian inference of parameters are absent. Although these topics are fairly new to population genetics, empirical investigators who wish to answer the question posed in the Preface—"given a collection of DNA sequences, what underlying forces are responsible for the observed patterns of variability?"—would be interested to know about them. Finally, although examples are used effectively to describe individual methods, some synthesis that connects several techniques to the same example dataset would help empirical population geneticists design analyses of their own data.

Despite these limitations, if used in conjunction with texts that state results in broader biological context, this book would be helpful to those with a mathematical or statistical background who are encountering the subject for the first time. The book is more theoretical than most books in the area, and because it is more modern than related classics, such as those by Crow and Kimura (1970) and Ewens (1979), it will be an informative reference for researchers in the field.

Noah A. ROSENBERG
University of Southern California

REFERENCES

- Crow, J. F., and Kimura, M. (1970), *An Introduction to Population Genetics Theory*, Minneapolis: Burgess.
- Ewens, W. J. (1979), *Mathematical Population Genetics*, Berlin: Springer-Verlag.
- Hartl, D. L., and Clark, A. G. (1998), *Principles of Population Genetics* (3rd ed.), Sunderland, MA: Sinauer.
- Li, W.-H. (1997), *Molecular Evolution*, Sunderland, MA: Sinauer.